Tensors: a Brief Introduction

Pierre Comon

Tensors: a Brief Introduction

Pierre Comon⋆, *Fellow, IEEE*

***Abstract*—Tensor decompositions are at the core of many Blind Source Separation (BSS) algorithms, either explicitly or implicitly. In particular, the Canonical Polyadic (CP) tensor decomposition plays a central role in identification of under-determined mixtures. Despite some similarities, CP and Singular Value Decomposition (SVD) are quite different. More generally, tensors and matrices enjoy different properties, as pointed out in this brief introduction.**

I. Motivation

Originally, Blind Source Separation (BSS) exploited mutual statistical independence between sources [20]. Among possible approaches based on the sole hypothesis of source statistical independence, several use cumulants. In fact, when random variables are independent, their cumulant tensor is diagonal [57]. When the source mixture is linear, the decomposition of the data cumulant tensor into a sum of outer products yields the columns of the mixing matrix. This is the first instance of tensor decomposition applied to BSS, even if it is not always explicit. In that case, the tensor is actually symmetric. In the presence of noise, the extraction of sources themselves needs another procedure, based for instance on a spatial matched filter (SMF) [20].

BSS has then been addressed later in different manners. A quite interesting class of approaches consists of exploiting an additional *diversity* [74]. More precisely, measurements are usually made in two dimensions, generally space and time. But if they are made as a function of three (or more) dimensions, *e.g.* frequency, polarization, time repetition, etc, the data are stored in a multi-way array. By treating this array as a matrix, information is lost. Yet, in some real-world applications, it is meaningful to assume a multi-linear model for this multi-way array, which justifies to consider it as a tensor. The decomposition of the latter into a sum of outer products yields not only the columns of the mixture, but also an estimate of the sources. So contrary to the first generation of BSS algorithms, there is no need to resort to an extracting filter. In addition, no statistics are to be estimated, so that the performance is expected to be better for short samples or correlated sources.

Beside numerous books dedicated to applications in physics, there already exist some surveys that can be used in Signal Processing. To start with, some background is presented in [46], *i.e.* basic engineering tools and a good panel of applications; a more signal processing oriented tensor overview may be found in [16]. A quite complete digest, more theoretical and oriented towards algebraic geometry, can be found in [49].

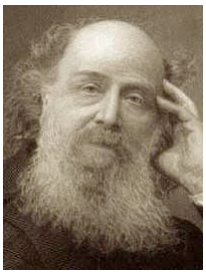
⋆ *Corresponding Author*: P. Comon, GIPSA-Lab, CNRS UMR5216, Grenoble Campus, BP.46, F-38402 St Martin d'Hères cedex, France. pierre.comon @ grenoble-inp.fr, tel: +33-4 7682 6271, fax: +33-4 7657 4790.
This work has been funded by the ERC Grant, Decoda project, agreement no. 320594.

The present survey aims at motivating the Signal Processing readership in diving in the promising world of tensors.

II. The world of tensors

Tensors have been introduced at the end of the nineteenth century with the development of the differential calculus. They have then been omnipresent in physics, to express laws independently of coordinate systems. Yet, a tensor is essentially a mapping from a linear space to another, whose coordinates transform *multilinearly* under a change of bases, as subsequently detailed. For an easier reading, we shall resort to arrays of coordinates, when this indeed eases presentation; interested readers may want to refer to [23], [49] for a more advanced coordinate-free presentation.

E. Waring
(1736-1798)



J. J. Sylvester
(1814 - 1897)

A. Clebsch
(1833-1872)

A. Linearity

Linearity expresses the property of a map μ defined on a vector space $\mathcal{S}$ onto another vector space $\mathcal{S}'$ built on the same field[1] $\mathbb{K}$ that: $\mu(\alpha\mathbf{x}+\beta\mathbf{y}) = \alpha\mu(\mathbf{x})+\beta\mu(\mathbf{y}), \forall \mathbf{x},\mathbf{y} \in \mathcal{S}, \alpha,\beta \in \mathbb{K}$. If $\mathcal{S}$ and $\mathcal{S}'$ are of finite dimension, then this map can be represented by a matrix of coordinates, once the bases of $\mathcal{S}$ and $\mathcal{S}'$ have been fixed. We see that every linear map can be associated with a matrix, say $\mathbf{A}$, so that $\mu(\mathbf{x}) = \mathbf{A}\mathbf{x}$. On the other hand, every matrix does not uniquely define a map. In fact, a matrix $\mathbf{A}$ could for instance define a bilinear form from $\mathcal{S} \times \mathcal{S}'$ onto $\mathbb{K}$, *i.e* $f(\mathbf{x}_1,\mathbf{x}_2) = \mathbf{x}_1^{\mathsf{T}}\mathbf{A}\mathbf{x}_2$. Hence, the correspondence between maps and arrays of coordinates is not one-to-one.

B. Bilinearity

Let's start with a simple example.

> *Example 1:* Consider two multi-dimensional zero-mean random variables $\mathbf{z}_1$ and $\mathbf{z}_2$, and denote the cross-covariance matrix by $\boldsymbol{\mathcal{G}} = \mathbb{E}\{\mathbf{z}_1\,\mathbf{z}_2^{\mathsf{T}}\}$. We see that the covariance is linear with respect to $\mathbf{z}_1$ and $\mathbf{z}_2$, which is referred to as *bilinearity*. Now suppose that $\mathbf{z}_1$ and $\mathbf{z}_2$ represent two phenomena that are

[1] As far as we are concerned, $\mathbb{K}$ will be either the field of real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$.

measured in a given coordinate system. $\boldsymbol{\mathcal{G}}$ gives an indication of their correlation. If we change the coordinate system, the covariance matrix changes. More precisely, if $\mathbf{z}'_1 = \mathbf{A}\mathbf{z}_1$ and $\mathbf{z}'_2 = \mathbf{B}\mathbf{z}_2$, then $\boldsymbol{\mathcal{G}}' = \mathbb{E}\{\mathbf{z}'_1 \mathbf{z}'^{\mathsf{T}}_2\}$ can be written $\boldsymbol{\mathcal{G}}' = \mathbf{A}\boldsymbol{\mathcal{G}}\mathbf{B}^{\mathsf{T}}$. We see that $\boldsymbol{\mathcal{G}}' \neq \boldsymbol{\mathcal{G}}$ whereas the phenomena remain the same. So we must distinguish between the physical phenomena that are coordinate-free, and the arrays of measurements we made. And because of bilinearity, we know how to go from one matrix representation to another. We may say that the covariance object is a *tensor* of order 2, and can be represented by a matrix in any given coordinate system.

What we just saw in Example 1 can be put in more formal terms. Now assume a linear change of coordinates is made in spaces $\mathcal{S}$ and $\mathcal{S}'$ defined by matrices $\{\mathbf{A}, \mathbf{B}\}$ so that the new coordinates express as $\mathbf{x}'_1 = \mathbf{A}\mathbf{x}_1$ and $\mathbf{x}'_2 = \mathbf{B}\mathbf{x}_2$. A tensor $\mathcal{G}$ represented in the original basis with an array $\boldsymbol{\mathcal{G}}$ will be represented (as in Example 1) in the new basis by the new array $\boldsymbol{\mathcal{G}}'$ whose coordinates are:

$$G'_{ij} = \sum_{p,q} A_{ip} B_{jq} G_{pq}$$

This can be compactly denoted by[2]: $\boldsymbol{\mathcal{G}}' = (\mathbf{A}, \mathbf{B}) \cdot \boldsymbol{\mathcal{G}}$. This will now be extended to orders higher than 2.

C. Multilinearity

Now assume $\mathcal{S}_d$ are D vector spaces, $1 \leq d \leq D$, and suppose f is a map from $\mathcal{S}_1 \times \cdots \times \mathcal{S}_D$ onto $\mathbb{K}$. Map f is said to be multilinear if $f(\mathbf{x}_1, \ldots, \mathbf{x}_D)$ is linear with respect to every variable x_d, $1 \leq d \leq D$. In other words, $f(\mathbf{x}_1, \ldots, \alpha\mathbf{x}_d + \beta\mathbf{y}_d, \ldots, \mathbf{x}_D) = \alpha f(\mathbf{x}_1, \ldots, \mathbf{x}_d, \ldots, \mathbf{x}_D) + \beta f(\mathbf{x}_1, \ldots, \mathbf{y}_d, \ldots, \mathbf{x}_D)$, $\forall d$, $\forall \alpha, \beta \in \mathbb{K}$. This map is actually a *multilinear form*. As in the previous section, map f can be represented by an array of coordinates, once the bases of $\mathcal{S}_d$ have been fixed, $1 \leq d \leq D$, and this array needs D indices.

D. Tensors

For the sake of simplicity, let us focus on $D = 3$, which is sufficient to give an idea. Because of multilinearity, special properties are satisfied. For instance $f(\alpha\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = f(\mathbf{x}_1, \mathbf{x}_2, \alpha\mathbf{x}_3)$, so that the two triplets of vectors $(\alpha\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ and $(\mathbf{x}_1, \mathbf{x}_2, \alpha\mathbf{x}_3)$ have the same image. When dealing with multilinear forms, it is hence relevant to consider the equivalence classes defined by the relation $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \sim (\mathbf{x}', \mathbf{y}', \mathbf{z}')$ if there exist $\alpha, \beta, \gamma \in \mathbb{K}$ such that $(\mathbf{x}', \mathbf{y}', \mathbf{z}') = (\alpha\mathbf{x}, \beta\mathbf{y}, \gamma\mathbf{z})$, with $\alpha\beta\gamma = 1$. Each class may be regarded as a *decomposable*[3] *tensor*. The space spanned by these classes is denoted as $\mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \mathcal{S}_3$, where $\otimes$ is called the *tensor product*. An element of this space is called a *tensor of order*[4] *3*. In more mathematical words, one would say that $\mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \mathcal{S}_3$ is the quotient space $\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3 / \sim$.

[2]Another notation, equally acceptable, is sometimes used: $\boldsymbol{\mathcal{G}}' = [\![\boldsymbol{\mathcal{G}}; \mathbf{A}, \mathbf{B}]\!]$.

[3]Decomposable tensors are also called pure or simple.

[4]In physics, the word *rank* is also sometimes used, but we shall avoid it because of the possible confusion with the more standard meaning related to *rank* of a linear operator.

Example 2: Let $\mathbf{x}_1 \in \mathcal{S}_1$, $\mathbf{x}_2 \in \mathcal{S}_2$ and $\mathbf{x}_3 \in \mathcal{S}_3$. Tensors $6\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3$ and $\mathbf{x}_1 \otimes 2\mathbf{x}_2 \otimes 3\mathbf{x}_3$ are the same, but in $\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3$, vectors $(6\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ and $(\mathbf{x}_1, 2\mathbf{x}_2, 3\mathbf{x}_3)$ are different.

If a linear change of basis is made in space $\mathcal{S}_1$ (resp. $\mathcal{S}_2$ and $\mathcal{S}_3$), as $\mathbf{x}' = \mathbf{A}\mathbf{x}$ (resp. $\mathbf{y}' = \mathbf{B}\mathbf{y}$ and $\mathbf{z}' = \mathbf{C}\mathbf{z}$), then the array $\boldsymbol{\mathcal{T}}'$ defining multilinear form f in the new coordinate system expresses as a function of $\boldsymbol{\mathcal{T}}$. For so-called *contravariant* tensors, the relationship is

$$T'_{ijk} = \sum_{pqr} A_{ip} B_{jq} C_{kr}\, T_{pqr} \tag{1}$$

as in Example 1, or in compact form: $\boldsymbol{\mathcal{T}}' = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{\mathcal{T}}$. On the other hand, there also exist *covariant* tensors for which the inverses of the above matrices are instead involved (cf. Example 4), and even mixed tensors that are partly covariant and partly contravariant [71], [23]. However, we shall concentrate only on contravariant tensors in this paper, which follow (1) under a multilinear transformation. Note that (1) holds true for contravariant tensors even if the linear transforms $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ are not invertible; they can even be rectangular matrices. This property is crucial in BSS when mixtures are underdetermined [20], [83].

Example 3: Consider three multi-dimensional random variables $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Then the 3rd order moment tensor $\mathcal{M}$ is represented by the 3rd order array $M_{ijk} = \mathbb{E}\{x_i y_j z_k\}$. As in the case of 2nd order moments, it is a contravariant tensor. In fact, if $\mathbf{x}' = \mathbf{A}\mathbf{x}$, $\mathbf{y}' = \mathbf{B}\mathbf{y}$ and $\mathbf{z}' = \mathbf{C}\mathbf{z}$, then $\boldsymbol{\mathcal{M}}' = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{\mathcal{M}}$ as in (1). It turns out that cumulants may also be seen as tensors as pointed out in [57]. Because cross-cumulants of independent random variables are null at any order, they have been extensively used in BSS. For instance, the cumulant tensor of order 2 is nothing else but the covariance matrix, and accounts for the correlation at order 2 only; it is not sufficient to account for statistical independence unless variables are Gaussian.

Example 4: The derivatives of order D of a multivariate scalar function can be stored in a covariant tensor of order D. For instance at order 2, if $A_{ij} = \partial^2 f / \partial x_i \partial x_j$, $\mathbf{x}' = \mathbf{M}\mathbf{x}$ and $A'_{pq} = \partial^2 f / \partial x'_p \partial x'_q$, then $\mathbf{A}' = \mathbf{N}^{\mathsf{T}}\mathbf{A}\mathbf{N}$, with $\mathbf{N} = \mathbf{M}^{-1}$. From now on and for the sake of simplicity, we shall only consider contravariant tensors in this paper.

More generally, a tensor of order D is an element of $\mathcal{S}_1 \otimes \ldots \otimes \mathcal{S}_D$, and can be represented by a D-way array $\boldsymbol{\mathcal{T}}$ once bases of spaces $\mathcal{S}_d$ have been fixed. Under multilinear transforms, these arrays of coordinates change similarly to (1).

Example 5: In physics, Hooke's law relates the deformation (strain) of a solid under the action of forces (stress). It states that stress $\mathcal{F}$ is related to strain $\mathcal{X}$ by the elasticity tensor as: $\mathcal{F} = \mathcal{C} \bullet \mathcal{X}$, where $\bullet$ is a contraction operator (see Section II-F for a formal definition). Once bases are fixed in

the stress and strain spaces, this relationship can be written in terms of arrays of coordinates:

$$F_{ij} = \sum_{p,q} C_{ijpq}\, X_{pq}$$

The elasticity tensor $\mathcal{C}$ is of order 4. Strain and stress are tensors of order 2, which are represented by matrices.

As illustrated above, it should be kept in mind that an array of coordinates alone does not suffice to define a tensor: spaces and bases need to be defined. Since we are interested mainly in manipulating arrays, and not so much in the map they may represent, arrays will be subsequently associated with *multilinear forms*, that is, maps from a product of spaces to their construction field $\mathbb{K}$. Even if most results can be stated without introducing arrays of coordinates [49], bases are required in engineering applications because calculations are made with arrays of numbers.

E. Notation

In the literature, indices of D-way arrays are sometimes put in superscripts or in subscripts, depending on the covariant or contravariant character of corresponding subspaces; this notation also allows the use the Einstein summation convention. Because we consider essentially fully contravariant tensors in this paper, we do not need to make the distinction.

Throughout the paper, arrays of numbers will be printed in boldface. More precisely, one-way and two-way arrays will be denoted in bold lowercase and bold uppercase, respectively, like for instance $\mathbf{v}$ and $\mathbf{M}$. Arrays with more than two indices will be denoted by bold calligraphic symbols, as $\boldsymbol{\mathcal{A}}$. Sets and Spaces will be noted in script font, like $\mathcal{S}$, whereas tensors will be printed in calligraphic font, as $\mathcal{A}$. Entries of arrays $\mathbf{v}$, $\mathbf{M}$ and $\boldsymbol{\mathcal{A}}$ will be noted v_i, M_{ij} and $A_{ij..k}$, without bold font because they are scalar numbers. In practice, a tensor $\mathcal{A}$ is often assimilated to its array representation $\boldsymbol{\mathcal{A}}$ [46], [16], [21], which is generally not so much confusing. Nevertheless, we shall make the distinction in the sequel, to keep the presentation as clear as possible.

F. Transformations

The tensor product $\mathcal{A} \otimes \mathcal{B}$ between two tensors $\mathcal{A} \in \mathcal{S}_1 \otimes \mathcal{S}_2$ and $\mathcal{B} \in \mathcal{S}_3 \otimes \mathcal{S}_4$ is a tensor of $\mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \mathcal{S}_3 \otimes \mathcal{S}_4$. The consequence is that the orders add up under tensor product.

Example 6: Let $\mathcal{A}$ be represented by a 3-way array $\boldsymbol{\mathcal{A}} = [A_{ijk}]$ and $\mathcal{B}$ by a four-way array $\boldsymbol{\mathcal{B}} = [B_{\ell mnp}]$; then tensor $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$ is represented by the 7-way array of components $C_{ijk\ell mnp} = A_{ijk} B_{\ell mnp}$. With some abuse of notation, the tensor product is often applied to arrays of coordinates, so that notation $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} \otimes \boldsymbol{\mathcal{B}}$ may be encountered.

If the tensor product increases the order, the *contraction* decreases it by 2. Contraction consists in a summation over a pair of indices. This operation permits to define the mode-k product between tensors, and can be denoted by $\bullet_k$, where k indicates which index should be summed.

Example 7: If $\mathcal{A}$ and $\mathcal{A}'$ are tensors of order D and D', the tensor $\mathcal{B} = \mathcal{A} \bullet_k \mathcal{A}'$ is a tensor of order $D + D' - 2$ obtained by summing over the kth index. For instance if $(D, D', k) = (3, 3, 2)$, this yields $B_{ijpq} = \sum_\ell A_{i\ell j} A'_{p\ell q}$. For $(D, D', k) = (2, 2, 1)$, we would have the matrix product $\boldsymbol{\mathcal{A}} \bullet_1 \boldsymbol{\mathcal{A}}' = \mathbf{A}^\mathsf{T} \mathbf{A}'$. However, when the product is between a matrix and a tensor of higher order, it has been the usual practice to always sum over the second matrix index. For instance if $\mathbf{M}$ is a matrix, $\mathcal{A} \bullet_3 \mathbf{M}$ means that the sum is performed on the 3rd tensor index and the 2nd matrix index.

It may be convenient to store D-way arrays in matrices. This transformation is called matrix *unfolding* or *flattening*, and can be performed in different manners, depending on the arbitrarily chosen ordering [46], [27]. Here, the ordering of [46] has been retained, but the choice of [27] would work equally well. In fact, the exact definition is not so important, provided the inverse map is defined consistently. We shall limit ourselves to matrices whose number of rows equals one of the tensor dimensions; this is sometimes referred to as *mode-n unfolding* [46]. Example 8 illustrates how to relate a 3rd order tensor to its 3 flattening matrices. But it is also possible to associate a D-way array, $D > 3$, to a multilinear operator of lower order, see *e.g.* [29], [64], [69], [9].

Example 8: Let a $2 \times 2 \times 2$ array of coordinates A_{ijk}. Its mode-n unfoldings $\mathbf{A}^{(n)}$ are:

$$\mathbf{A}^{(1)} = \left[\begin{array}{cc|cc} A_{\mathbf{1}11} & A_{\mathbf{1}21} & A_{\mathbf{1}12} & A_{\mathbf{1}22} \\ A_{\mathbf{2}11} & A_{\mathbf{2}21} & A_{\mathbf{2}12} & A_{\mathbf{2}22} \end{array}\right]$$

$$\mathbf{A}^{(2)} = \left[\begin{array}{cc|cc} A_{1\mathbf{1}1} & A_{2\mathbf{1}1} & A_{1\mathbf{1}2} & A_{2\mathbf{1}2} \\ A_{1\mathbf{2}1} & A_{2\mathbf{2}1} & A_{1\mathbf{2}2} & A_{2\mathbf{2}2} \end{array}\right]$$

$$\mathbf{A}^{(3)} = \left[\begin{array}{cc|cc} A_{11\mathbf{1}} & A_{21\mathbf{1}} & A_{12\mathbf{1}} & A_{22\mathbf{1}} \\ A_{11\mathbf{2}} & A_{21\mathbf{2}} & A_{12\mathbf{2}} & A_{22\mathbf{2}} \end{array}\right]$$

Remark that the row number of matrix $\mathbf{A}^{(n)}$ corresponds to the nth index of tensor $\mathcal{A}$.

The Kronecker product between two matrices $\mathbf{A}$ and $\mathbf{B}$ of size $I \times J$ and $K \times L$, respectively, is the matrix $\mathbf{A} \boxtimes \mathbf{B}$ of size $IK \times JL$ defined blockwise by

$$\mathbf{A} \boxtimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & \dots & A_{1J}\mathbf{B} \\ \vdots & & \vdots \\ A_{I1}\mathbf{B} & \dots & A_{IJ}\mathbf{B} \end{bmatrix}.$$

The Kronecker product is used to represent the tensor product when bases are fixed, and when tensors are represented by their array of coordinates unfolded into matrices. It should be borne in mind that the Kronecker product usually applies to matrices (although an extended definition has recently been proposed in [63]), whereas the tensor product is more general and coordinate-free. Hence they should not be confused.

G. Special tensors

A particularly important class of tensors is that of *decomposable* tensors, which are tensor products of vectors. As already said in Section II-D, they are of the form $\mathcal{D} =$

$\mathbf{u} \otimes \mathbf{v} \otimes \ldots \otimes \mathbf{w}$, and span the whole tensor space. The corresponding array of coordinates is $D_{ij..k} = u_i v_j .. w_k$. One can view these tensors as a discretization of a multivariate function whose variables separate.

Example 9: Take a function of two variables with separated variables: $f(x,y) = u(x)v(y)$. Then its discretization takes the form $f(x_i, y_j) = u(x_i)v(y_j)$, and these numbers can be stored in a rank-one matrix $\mathbf{D} = \mathbf{u}\mathbf{v}^\mathsf{T}$.

A tensor is cubical[5] if all its spaces of construction are identical, with same basis. A cubical tensor $\mathcal{A}$ is *symmetric* if its array of coordinates is invariant under permutation of its indices: $A_{\sigma(ij..k)} = A_{ij..k}$, $\forall \sigma$.

Example 10: The tensor of moments and the tensor of derivatives, defined in Examples 1, 3 and 4, are symmetric.

The simplest symmetric array is the *diagonal* one, defined by $\Delta_{ij..k} = 0$ if $(i, j, .., k) \neq (i, i, .., i)$.

III. Decompositions and ranks

A. Tensor rank

Any tensor $\mathcal{T}$ can be decomposed (non uniquely) into a linear combination (with coefficients in $\mathbb{K}$) of decomposable tensors:

$$\mathcal{T} = \sum_{r=1}^{R} \lambda_r \, \mathcal{D}(r), \tag{2}$$

$\mathcal{D}(r) = \mathbf{a}_r \otimes \mathbf{b}_r \otimes \ldots \otimes \mathbf{c}_r$. If tensor spaces are endowed with scalar products, one can impose decomposable tensors $\mathcal{D}(r)$ to be built with unit norm vectors, which permits to impose $\lambda_r \in \mathbb{R}^+$ if desired. The smallest value R for which (2) holds is called the *tensor rank*. The definition of tensor rank can be traced back to the beginning of the 20th century [38], but it has been re-introduced in other disciplines under various names [12], [36], [39], [7], [82], [66].

Example 11: Let the arrays $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ of dimensions $2 \times 2 \times 2$ be defined by their mode-1 unfoldings:

$$\mathbf{A}^{(1)} = \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\mathbf{B}^{(1)} = \left[\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right]$$

Tensor $\mathcal{A} = [1,0] \otimes [1,0] \otimes [1,1]$ has rank 1. Tensor $\mathcal{B}$ is symmetric and has rank 3, as will be seen with $\mathcal{T}_0$ in Example 18.

Note that, by definition, a tensor is decomposable if and only if it has rank one. If the order of a tensor $\mathcal{T}$ is ≥ 3, the rank may depend on the field, in the sense that a real tensor of rank R may have smaller rank if we allow the decomposition (2) to be complex, as demonstrated in the example below.

Example 12: Take a real symmetric array $\boldsymbol{\mathcal{Y}}$ of dimensions $2 \times 2 \times 2$, defined by its mode-1 unfolding

$$\mathbf{Y}^{(1)} = \left[\begin{array}{cc|cc} 2 & 0 & 0 & -2 \\ 0 & -2 & -2 & 0 \end{array}\right]$$

Then, we need three decomposable tensors in $\mathbb{R}$:

$$\boldsymbol{\mathcal{Y}} = 4 \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} -1 \\ -1 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} -1 \\ 1 \end{bmatrix}^{\otimes 3}$$

but only two in $\mathbb{C}$, setting $\jmath = \sqrt{-1}$:

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} 1 \\ \jmath \end{bmatrix}^{\otimes 3} + \begin{bmatrix} 1 \\ -\jmath \end{bmatrix}^{\otimes 3}$$

Hence its tensor rank in $\mathbb{R}$ is 3 whereas it is 2 in $\mathbb{C}$.

Other examples may be found in [48], [18], [46]. Examples 11 and 12 incidentally show that, unlike matrix rank, tensor rank may exceed all dimensions.

B. Tucker decomposition

At this stage, it is interesting to make a connection with the matrix Singular Value decomposition (SVD). Two important features characterize the SVD of a matrix $\mathbf{M}$:

$$\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T} \tag{3}$$

namely (i) $\mathbf{U}$ and $\mathbf{V}$ have orthonormal columns, and (ii) $\boldsymbol{\Sigma}$ is diagonal. Consider the decomposition below of a three-way array, introduced by Tucker in the sixties [85]:

$$\mathcal{T}_{ijk} = \sum_p \sum_q \sum_r A_{ip} \, B_{jq} \, C_{kr} \, G_{pqr} \tag{4}$$

which we shall compactly denote $\boldsymbol{\mathcal{T}} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{\mathcal{G}}$. It is clear that if the number of free parameters in the right-hand side of (4) is smaller than the number of equations, then there will generally be no solution. This happens to be the case if $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are orthonormal and $\boldsymbol{\mathcal{G}}$ is diagonal. In the quest for existence, we have to choose: either $\boldsymbol{\mathcal{G}}$ is diagonal, but we have to relax the orthogonality constraint on factor matrices, which will be allowed to have more columns than rows (this corresponds to decomposition (2)), or we keep the orthonormality constraint, but allow $\boldsymbol{\mathcal{G}}$ to have nonzero extra-diagonal entries as elaborated in the next section.

C. HOSVD and multilinear ranks

If we impose matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ to have orthogonal and unit-norm columns in the Tucker decomposition (4), then we can make several observations. First, denote by R_n the rank of $\mathbf{T}^{(n)}$, the nth unfolding matrix of $\boldsymbol{\mathcal{T}}$, $1 \leq n \leq D = 3$. Rank R_n is called *mode-n rank* of $\mathcal{T}$, or n-rank in short. Then the number of columns of $\mathbf{A}$ (resp. $\mathbf{B}$, $\mathbf{C}$) does not need to exceed R_1 (resp. R_2, R_3), and the dimension of the core tensor may be imposed to be $R_1 \times R_2 \times R_3$. In addition R_n cannot exceed[6] the tensor rank R defined in (2), nor the nth dimension. The D-uple of n-ranks is the *multilinear rank* of $\mathcal{T}$. Another property is less immediate to capture: the core array $\boldsymbol{\mathcal{G}}$ can be imposed to be *all-orthogonal*, which means that all tensor slices of order $D-1$ are orthogonal to each other in every mode; when $D = 3$ this means:

$$\sum_{j,k} G_{\alpha jk} G_{\beta jk} = \sum_{i,k} G_{i\alpha k} G_{i\beta k} = \sum_{i,j} G_{ij\alpha} G_{ij\beta} = 0$$

[5] the terminology of *homogeneous* is also used in physics.

[6] This property is not a surprise, if we view decomposition (2) as a decomposition of the nth unfolding matrix into a sum of rank-1 matrices where rows are imposed to have a special structure.

if $\alpha \neq \beta$. See [27] and references therein for more details. It is worth to notice the elementary fact that for tensors of order 2 (*i.e* matrices), $R_1 = R_2 = R$, and all equal the matrix rank.

Example 13: The multilinear rank of array $\boldsymbol{\mathcal{B}}$ defined in Example 11 is $(2, 2, 2)$, whereas that of $\boldsymbol{\mathcal{A}}$ is $(1, 1, 1)$.

D. CP decomposition

On the contrary, if we keep a diagonal form for $\boldsymbol{\mathcal{G}}$, we end up with the *polyadic decomposition* [38], also sometimes called Candecomp or Parafac because of its rediscovery in the seventies:

$$\mathcal{T}_{ijk} = \sum_{r=1}^{R} A_{ir} B_{jr} C_{kr} \, \lambda_r \tag{5}$$

or in compact form $\boldsymbol{\mathcal{T}} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \boldsymbol{\mathcal{L}}$, where $\boldsymbol{\mathcal{L}}$ is diagonal. If R is not loo large, this decomposition can be unique (cf. Section V) and deserves to be called *Canonical Polyadic* (CP). Following a practice now adopted in applied mathematics and engineering [42], [5], we shall subsequently use the acronym CP, which can also cleverly stand for Candecomp/Parafac. After inspection, it may be seen that (5) is nothing else but decomposition (2) in array coordinates. In other words, the CP decomposition reveals the tensor rank.

E. Symmetric rank

As already pointed out in Section II-G, a tensor $\mathcal{T}$ is symmetric if its coordinate array $\boldsymbol{\mathcal{T}}$ is invariant by permutations of indices. If we impose tensors $\mathcal{D}(r)$ in (2) to be themselves symmetric, then we might end up with a larger value of rank, denoted R_s, which is referred to as the *symmetric rank* of $\mathcal{T}$. It is clear that $R_s \geq R$ for any symmetric tensor $\mathcal{T}$, since any constraint on decomposable tensors may increase rank; we have already observed this fact with the real constraint in Example 12. It has been conjectured in [19] that rank and symmetric rank are always equal, but this has not yet been proved in the general case.

F. Nonnegative rank

When an array is real nonnegative, one may want to impose rank-1 terms in its CP decomposition to be themselves nonnegative. The minimal number of terms is then called the *nonnegative rank* and is generally strictly larger than the rank in $\mathbb{R}$. This is already the case for matrices ($D = 2$) as shown in the example below, due to Herbert E. Robbins. The same phenomenon is observed for tensors, although theoretical results are still lacking.

Example 14: The following matrix has rank 3 since vector $[1, -1, -1, 1]$ belongs to its kernel. But it can be proved that its nonnegative rank is 4.

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

G. Structured ranks

More generally, when matrix factors are imposed to have a special structure, such as banded, van der Monde, Toepltiz or Hankel, the tensor rank may increase, just as in the nonnegative case. Structure can also have an impact on computational issues [49], [78].

H. Border rank

A tensor has border rank $\underline{R}$ if it is the limit of tensors of rank $\underline{R}$ and not the limit of tensors of smaller rank. Rank and border rank always coincide for matrices, but not for tensors of order larger than 2, as shown in the next example.

Example 15: Let u and v be fixed real or complex numbers, and ε a small positive number. Then $\frac{1}{\varepsilon}\left[(u + \varepsilon v)^3 - u^3\right] = 3\,u^2 v + O(\varepsilon)$. Now if multiplication is not commutative, we have 3 distinct terms in the right hand side; this is what happens for the tensor product, so that $\forall \varepsilon > 0$:

$$\begin{aligned} \mathcal{T}_\varepsilon &= \frac{1}{\varepsilon}\left[(\mathbf{u} + \varepsilon \mathbf{v})^{\otimes 3} - \mathbf{u}^{\otimes 3}\right] = \mathcal{T}_0 + O(\varepsilon), \\ \mathcal{T}_0 &= \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u} \end{aligned}$$

hold for any vectors $\mathbf{u}$ and $\mathbf{v}$. If the latter are not collinear, it can be proved that $\mathcal{T}_0$ is of rank $R = 3$, but is the limit of tensors $\mathcal{T}_\varepsilon$, which are all of rank 2. Hence the border rank of $\mathcal{T}_0$ is $\underline{R} = 2$.

The border rank has been defined and utilized by many authors, especially in arithmetic complexity [7], [72], [82], [52]. This concept is crucial in tensor approximation problems, as addressed in Section VI.

IV. Relation with polynomials

Homogeneous polynomials are bijectively related to tensors, which allows to transpose existing results of algebraic geometry; see *e.g.* [49], [15], [19], [80], [10], [23], [17] and references therein. In fact, one can associate the following polynomial with any array $\boldsymbol{\mathcal{T}}$:

$$p(\mathbf{x}, \mathbf{y}, \ldots \mathbf{z}) = \sum_{i,j,..k} T_{ij..k} \, x_i y_j \ldots z_k$$

Conversely, any homogeneous polynomial of degree D and partial degree 1 in every variable can be associated with a (non symmetric) tensor $\mathcal{T}$.

Through this bijection, a decomposable tensor of order D is translated into a product of D linear forms, and the CP decomposition can be translated into a linear combination of such terms:

$$p(\mathbf{x}, \mathbf{y}, \ldots \mathbf{z}) = \sum_{r=1}^{R} \lambda_r \, (\mathbf{a}_r^{\mathsf{T}} \mathbf{x})(\mathbf{b}_r^{\mathsf{T}} \mathbf{y}) \ldots (\mathbf{c}_r^{\mathsf{T}} \mathbf{z}) \tag{6}$$

In the case of symmetric tensors, $\mathbf{x} = \mathbf{y} = \cdots = \mathbf{z}$. More precisely, a symmetric tensor $\mathcal{T}$ of order D can be identified with the homogeneous polynomial of degree D:

$$p(\mathbf{x}) = \sum_{i,j,..k} T_{ij..k} \, x_i x_j \ldots x_k$$

in the indeterminates $x_1, \ldots, x_n$. It can be easily checked that symmetric tensors of rank one are associated with a polynomial of the form: $a(\mathbf{x})^D$, where $a(\mathbf{x}) = \mathbf{a}^\mathsf{T}\mathbf{x}$ is a linear form. In other words, they are exactly the D^{th} powers of a homogeneous linear form. The CP decomposition of $\mathcal{T}$ reduces in this case to:

$$p(\mathbf{x}) = \sum_{i=1}^{R_s} a_i(\mathbf{x})^D \tag{7}$$

which has been classically called a Waring decomposition [40]. The minimum number of summands R_s in a Waring decomposition is the symmetric rank of $\mathcal{T}$, which we have defined earlier.

Example 16: The polynomials associated with tensors $\mathcal{A}$ and $\mathcal{B}$ of Example 11 are respectively: $a(x_1, x_2, y_1, y_2, z_1, z_2) = x_1 y_1 (z_1 + z_2)$ and $b(x_1, x_2) = 3x_1^2 x_2$.

Example 17: Take the polynomial of degree $D = 3$:

$$\begin{aligned} 2x_1^3 - 6x_1x_2^2 &= (x_1 + \jmath x_2)^3 + (x_1 - \jmath x_2)^3 \\ &= 4(x_1)^3 - (x_1 + x_2)^3 - (x_1 - x_2)^3 \end{aligned}$$

where $\jmath = \sqrt{-1}$. It has complex symmetric rank equal to 2 and real symmetric rank equal to 3. This polynomial is actually associated with tensor $\mathcal{Y}$ given in Example 12.

Example 18: Example 15 can be written in terms of polynomials, and is even easier to understand this way. Take $\mathbf{u} = [1, 0]$ and $\mathbf{v} = [0, 1]$. Then $\mathbf{u}^{\otimes 3}$ and $\mathbf{v}^{\otimes 3}$ are associated with polynomials x_1^3 and x_2^3 respectively, whereas $(\mathbf{u} + \varepsilon \mathbf{v})^{\otimes 3}$ is associated with $(x_1 + \varepsilon\, x_2)^3$, which can be expanded as $x_1^3 + 3\varepsilon\, x_1^2 x_2 + o(\varepsilon)$. This shows that $\mathcal{T}_\varepsilon$ is associated with $3x_1^2 x_2 + o(\varepsilon)$. Hence $\mathcal{T}_\varepsilon$ tends to $\mathcal{T}_0$, because $\mathcal{T}_0$ is associated with $3x_1^2 x_2$. Moreover, the rank of $\mathcal{T}_0$ is 3 because $3x_1^2 x_2$ cannot be written as a sum of fewer than 3 cubes.

V. Exact decompositions

Now one can ask oneself the question whether the CP decomposition defined in (2) and (5) is unique or not. First of all, the D-way array associated with a Dth order decomposable tensor $\mathcal{D}$ is not uniquely represented by an outer product of vectors: there remain $D - 1$ scaling factors of unit modulus. So we are rather interested in the uniqueness of coefficients $\lambda_r \in \mathbb{R}^+$ and tensors $\mathcal{D}(r)$, which is more meaningful; this is sometimes called *essential uniqueness*. We shall see in this section two ways of assessing uniqueness: almost surely or deterministically.

A. Expected rank

A naive approach is to count the number of degrees of freedom on both sides of (6), which is a rewriting of (2) in terms of polynomials, and say that the number of equations should be at least as large as the number of unknowns. To fix the ideas, take a tensor of order D and dimensions $n_1 \times \cdots \times n_D$. It is clear that a necessary condition for uniqueness of the CP decomposition is that

$$R\left((\sum_{i=1}^{D} n_i) - D + 1\right) \leq N \tag{8}$$

where $N = \prod_i n_i$. We can proceed similarly for symmetric tensors and count equations and unknowns in (7). This leads to

$$R_s\, n \leq N_s \tag{9}$$

where $N_s = \binom{n+D-1}{D}$ corresponds to the number of free parameters in a symmetric tensor. Equations (8) and (9) induce an upper bound on rank, which is called the *expected rank*, and is defined as

$$R \leq R^o = \left\lceil \frac{N}{1 - D + \sum_i n_i} \right\rceil \tag{10}$$

$$R_s \leq R_s^o = \left\lceil \frac{N_s}{n} \right\rceil \tag{11}$$

When the fraction above is not an integer, there will always be an infinity of solutions, because of too many free parameters. When it is an integer, the number of unknowns is equal to the number of equations, and we could expect that there is a finite number of solutions. However, things are not so simple, as early pointed out by Clebsch in the nineteenth century. In fact, there are exceptions [3], [1], [23], [58].

Example 19: Consider 4th order symmetric tensors of dimension 3. In that case, $N_s/n = \binom{6}{4}/3 = 5$ is an integer. Our hope is disappointed, since 5 forms are generally not sufficient in their decomposition. This exception was first noticed by Clebsch from the polynomial framework: the "generic rank" of ternary quartics is in fact 6 [33]. This means that most homogeneous polynomials of degree 4 in 3 variables in $\mathbb{C}$ can be written as a sum of 6 linear forms raised to the 4th power, and not fewer with probability 1.

B. Typical and generic ranks

Generic (resp. typical) ranks are the ranks that we encounter with probability one (resp. nonzero probability), when their entries are drawn independently according to a continuous probability distribution, hence their importance. Contrary to the matrix case, they are not maximal; tables of rank values may be found in [24], as well as simple codes[7] to compute numerically the generic rank of a large panel of tensors.

A striking fact is that only one rank occurs with probability one (the so-called generic rank) in $\mathbb{C}$, whereas several typical ranks may exist in $\mathbb{R}$. The generic rank in $\mathbb{C}$ is always equal to the smallest typical rank one would find in $\mathbb{R}$. This problem was first addressed by Sylvester in XIX century. The case of real symmetric tensors of dimension 2 is now well understood [67], [22], [13]. In fact, all the integers between $\lfloor \frac{D+2}{2} \rfloor$ and D have been shown to be typical ranks [8]. If the tensor rank is

[7] Codes can be downloaded from the home page of the author.

smaller than a bound depending on the generic rank (typically R^o-1 as defined in equations 10-11), there exist almost surely finitely many CP decompositions. See [23] for a survey of recent results on almost sure uniqueness.

C. Uniqueness results based on linear algebra

Instead of associating tensors with polynomials and making use of results borrowed from algebraic geometry, uniqueness conditions can be obtained by considering particular factor matrices. However, these conditions are generally only sufficient [41], and often much more restrictive. The most well known is that published by Kruskal [47] and extended later in [73], [81]; alternate proofs have been derived in [68], [49]. It requires the following definition: The Kruskal rank of a matrix is the largest number κ such that *any subset* of κ columns is full rank. By construction, Kruskal's rank cannot exceed matrix rank.

Example 20: The matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

has rank 3, but its Kruskal rank is $\kappa = 2$.

The CP decomposition is unique if the sufficient condition holds:

$$2R + D - 1 \leq \sum_{d=1}^{D} \kappa_d \tag{12}$$

where κ_d denotes the Kruskal rank of the dth factor matrix in the CP decomposition. Further recent deterministic results may be found in [25], [31], [32]. These results do not need algebraic geometry but advanced linear algebra (*i.e.* compound matrices formed of minors). They are sometimes much more powerful than Kruskal's bound.

D. Exact computation

Space is lacking to describe various existing algorithms. However, we provide below some pointers to related literature, among many others. In [6], algorithms to compute the symmetric rank of symmetric tensors of small border rank are proposed. When the rank is small, the symmetric CP decomposition can be computed with the help of Sylvester's algorithm [10]; when it is not unique, one CP decomposition can still be delivered. In [60], approaches based on special eigenvector computations are proposed. Direct computation is proposed in [4] for $2 \times n \times n$ arrays.

When one tensor dimension is large compared to its rank and to other dimensions, it is possible to compute the CP decomposition via a joint congruent diagonalization of its matrix slices; this has been first proposed in [50] for two matrix slices. In the presence of errors with more than two slices, such a diagonalization becomes approximate [25] and needs more care (cf. next section). In a similar spirit, for low-rank tensors of order larger than 3, one can also decrease the order by working jointly on tensor slices of lower orders [29], or by rearranging the original tensor into another of lower order but larger dimensions [64].

VI. Approximate decompositions

In practice, measurements are always corrupted by some noise, which almost always has a continuous probability distribution. For this reason, the tensor rank is *generic* or *typical*, and the CP decomposition is generally not unique. That's why a best rank-r approximation must be computed [44] [21]. General-purpose optimization algorithms will generally suffice to solve the problem, *e.g.* [77], [65], [46], [21], [84]; they are widely used but their convergence towards a minimum is not guaranteed, because the objective function may have only an infimum.

In fact, low-rank approximations are useful and even unavoidable, but unfortunately *ill-posed* in general [37], [75], [75], except for special cases of tensors under constraints, like non negativity [54]. Most algorithms presently utilized by engineering communities ignore this fact, which may raise serious practical problems in a small fraction of cases.

Ill-posedness comes from the fact that the set of tensors of rank at most R is not *closed*, as pointed out in section III-H. Some remedies have been proposed in the literature to face or circumvent this difficulty. In practice, this means that *another problem* is solved, often by imposing constraints in the CP decomposition.

These include: (i) impose orthogonality between columns of factor matrices [20] - in Blind Source Separation, this takes the form of a spatial prewhitening; (ii) impose orthogonality between decomposable tensors [45]; (iii) prevent divergence by bounding coefficients λ_r [61], [54]; (iv) if the tensor is nonnegative, use a nonnegative CP [54]; (v) impose a minimal angle between columns of factor matrices [55]; (vi) compute an exact CP of another tensor[8], which has undergone a multilinear compression via truncated HOSVD [21], [11]; (vii) compute another decomposition where the core tensor is block diagonal instead of diagonal [26] [79]; (viii) compute a Joint Approximate Diagonalization (JAD) of matrix slices, which may be viewed as another decomposition where the core tensor is not diagonal [62], [87], [89], [2], [86], [51], [20], [30], [56], [69], [14], as depicted in Figure 1. The drawbacks of this family of approaches, which become more and more popular, are three-fold. First, rank must be smaller than two dimensions; in [25], the latter constraint is nevertheless relaxed. Second, replacing the core tensor by its diagonal yields an approximate CP decomposition whose optimality is not known. Third, a *closed* subclass of invertible matrices needs to be (arbitrarily) chosen, and indeed varies from one algorithm to another. (ix) When one dimension is much larger than the others, the optimality of this kind of approach can be significantly improved by imposing a structure in the diagonalization process [25].

Some codes are freely available on the internet. See for instance home pages of R. Bro, L. De Lathauwer, T. Kolda, A.H. Phan and P. Comon [90], [91], [92], [93], [94]. A good site to find applications and related references is the Three-Mode Company's maintained by P. Kroonenberg [95].

[8]It may happen that the problem remains ill-posed after this type of compression, because reducing the mode-ranks does not necessarily reduce tensor rank, even if it often does.

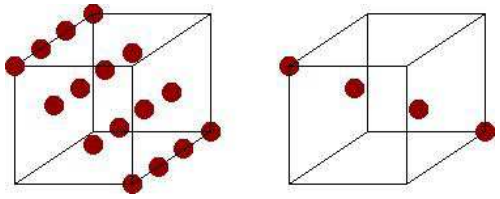


Fig. 1. Because the optimization criteria are different in JAD and CP decompositions, one does not attempt to zero the same entries. This figure shows the location of the entries that are not minimized in the core tensor (i) in the CP decomposition of a 3rd order $4 \times 4 \times 4$ tensor (right), and (ii) during the execution of a JAD algorithm (left). Note that JAD algorithms deliver two factor matrices; the entries of third one remain in the core tensor.

VII. The case of rank-one approximate

The rank-one approximation problem is of interest for at least two reasons: first it is always well-posed, and second it shows up in the deflation approach of BSS [20]. In addition, it is much easier to compute than a full CP decomposition [43] [28]. This problem may be seen to be related to tensor eigenvalues [17] [53] [59] [35] [88]. It has been proved recently that the best rank-1 approximation of a symmetric tensor is symmetric [34]; a shorter proof can be found in [35], as well as uniqueness issues. So a question deserves to be raised: can the exact or approximate CP decompositions be computed by successive rank-1 approximations? It is already known that this does not generally work.

In fact, attention should be paid to the fact that subtracting the best rank-1 approximate does not decrease tensor rank in general [80], contrary to the matrix case. Simple examples may be found in [18]; similar examples also exist for non symmetric or nonnegative tensors. The consequence is that the rank-1 terms appearing in the best rank-k tensor approximation *are not the same* for different values of k. Hence, it is not possible to compute a full CP decomposition by solving successive best rank-1 approximations, contrary to what has been claimed by some authors[9]. However, whether deflation works in special cases (such as structured CP decompositions) is still an open question.

Example 21: The tensor defined by its mode-1 unfolding

$$\mathbf{T} = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \end{array}\right]$$

is of rank 2. Its best rank-1 approximate is [80]:

$$\mathbf{Y} = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{array}\right]$$

And one checks out that the difference

$$\mathbf{T} - \mathbf{Y} = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

is of rank 3. In this example, deflation does not permit to decrease tensor rank.

[9]This procedure, called *deflation*, works in BSS for other reasons. In fact, BSS does not only reduce to a low-rank tensor approximation, but also includes a regression stage.

VIII. Applications

Applications of tensor decompositions (essentially CP) include arithmetic complexity, separation of variables, Blind Identification of linear mixtures, Blind Source Separation, Data Mining, Spectroscopy, Antenna Array Processing, Phylogenetics... Tucker and HOSVD have other application fields, in which uniqueness is not requested, like data compression. For reasons of space, we shall now detail only one application of the CP decomposition, namely fluorescence spectroscopy [76], for which very few theoretical results can apply, unfortunately. The reader is invited to consult *e.g.* [46], [16], [20] for pointers to other applications.

An optical excitation applied to a solution produces several effects, including Rayleigh and Raman diffusions, and Fluorescence. If the latter effect can be isolated, it may allow to accurately measure the relative concentrations of fluorescent solutes. In fact, at low concentrations and in the presence of R fluorescent solutes, the Beer-Lambert law can be linearized and takes the form:

$$\mathcal{T}(x,y,z) = T_o \sum_{\ell=1}^{R} \gamma_\ell(x)\epsilon_\ell(y)c_\ell(z)$$

where x, y and z denote the fluorescence emission wavelength, the excitation wavelength, and the sample number, respectively, $\mathcal{T}$ is the fluorescence intensity measured as a function of the latter variables, $\gamma_\ell(x)$ denotes fluorescence emission spectrum of the ℓth solute, $\epsilon_\ell(y)$ its absorbance spectrum (sometimes called excitation spectrum), and $c_\ell(z)$ its relative concentration. In practice, only a finite number of samples are available, and measurements are made on discrete values within a limited spectral range, so that variables x, y and z take a finite number of values. In other words, we deal with a CP decomposition of a finite 3-way array, often of rather large dimensions (several hundreds). The particularity of this CP decomposition is that $\mathcal{T}$ is real nonnegative, as well as all the terms involved in its CP decomposition. Hence, R is the *nonnegative rank* of $\mathcal{T}$. The good news are that (i) the best low-rank approximate always exists [54], and that (ii) there are simple efficient numerical algorithms available for its computation [70]. The bad news are that known uniqueness results, which we have reviewed in this paper, are not appropriate for *nonnegative* CP decompositions. For instance, if *nonnegative rank* is plugged in place of rank in (12), the obtained sufficient condition is more restrictive, and does not even guarantee that factor matrices are nonnegative. This is the subject of ongoing research.

Acknowledgment

This work has been funded by the European Research Council under European Community's Seventh Framework Programme FP7/2007-2013 grant agreement number 320594.

Author

Pierre Comon graduated in 1982, and received the Doctorate degree in 1985, both from the University of Grenoble, France. He later received the Habilitation to Lead Researches in 1995, from the University of Nice, France. He was with the ISL laboratory,

Stanford University, CA, in 1987. He spent 13 years in industry, first with Crouzet-Sextant, Valence, France, between 1982 and 1985, and then with Thomson Marconi, Sophia Antipolis, France, between 1988 and 1997. He joined in 1997 the Eurecom Institute at Sophia Antipolis and left during the fall of 1998. He is research director with CNRS since 1998, at laboratory I3S Sophia Antipolis until 2012, and since then at Gipsa-Lab, Grenoble. His research interests include High-Order Statistics (HOS), Blind Source Separation, Statistical Signal and Array Processing, Tensor decompositions, Multi-Way Factor Analysis and Data Mining, with applications to biomedical end environment.

Dr. Comon was Associate Editor of the *IEEE Transactions on Signal Processing* from 1995 to 1998, of the Eurasip journal *Signal Processing* from 2006 to 2011, and a member of the French *National Committee* of Scientific Research from 1995 to 2000. He was the coordinator of the European Basic Research Working Group on HOS, *ATHOS*, from 1992 to 1995. Between 1992 and 1998, he was a member of the *Technical and Scientific Council* of the Thomson Group (now Thalès). Between 2001 and 2004 he acted as launching Associate Editor with the *IEEE Transactions on Circuits and Systems I*, in the area of Blind Techniques. He has been a member of *SAM* and *SPTM* IEEE Technical Committees of the Signal Processing Society, and a member of the *BSP* TC of the Circuit and Systems Society. He is a member of the editorial board of the *SIAM Journal on Matrix Analysis and Applications* since January 2011.

REFERENCES

[1] H. Abo, G. Ottaviani, C. Peterson, "Induction for secant varieties of Segre varieties", *Trans. Amer. Math. Soc.*, vol. 361, pp. 767–792, 2009, arXiv:math/0607191.

[2] L. Albera, A. Ferreol, P. Chevalier, P. Comon, "ICAR, a tool for blind source separation using fourth order statistics only", *IEEE Trans. Sig. Proc.*, vol. 53, no. 10, pp. 3633–3643, Oct. 2005, hal-00743890.

[3] J. Alexander, A. Hirschowitz, "Polynomial interpolation in several variables", *J. Alg. Geom.*, vol. 4, no. 2, pp. 201–222, 1995.

[4] J. M. F. Ten Berge, "Kruskal's polynomial for $2 \times 2 \times 2$ arrays and a generalization to $2 \times n \times n$ arrays", *Psychometrika*, vol. 56, no. 4, pp. 631–636, 1991.

[5] J. M. F. Ten Berge, "Simplicity and typical rank of three-way arrays", *Jour. Chemometrics*, vol. 18, no. 1, pp. 17–21, 2004.

[6] A. Bernardi, A. Gimigliano, M. Ida, "Computing symmetric rank for symmetric tensors", *J. Symbolic Computation*, vol. 46, no. 1, pp. 34–53, 2011.

[7] D. Bini, M. Capovani, F. Romani, G. Lotti, "O(n2.77) complexity for nxn approximate matrix multiplication", *Inform. Process. Lett.*, vol. 8, no. 5, pp. 234–235, 1979.

[8] G. Blekherman, "Typical real ranks of binary forms", *Foundations of Computational Math.*, 2013.

[9] M. Boizard, G. Ginolhac, F. Pascal, S. Miron, P. Forster, "Numerical performance of a tensor MUSIC algorithm based on HOSVD", in *EUSIPCO*, Marrakech, Morocco, 13-15 Dec. 2013.

[10] J. Brachat, P. Comon, B. Mourrain, E. Tsigaridas, "Symmetric tensor decomposition", *Lin. Alg. Appl.*, vol. 433, no. 11/12, pp. 1851–1872, Dec. 2010, hal:inria-00355713, arXiv:0901.3706.

[11] R. Bro, C. A. Andersson, "Improving the speed of multiway algorithms. Part II: Compression", *Chemometrics and Intelligent Laboratory Systems*, vol. 42, no. 1-2, pp. 105–113, 1998.

[12] J. D. Carroll, J. J. Chang, "Analysis of individual differences in multidimensional scaling via n-way generalization of Eckart-Young decomposition", *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sept. 1970.

[13] A. Causa, R. Re, "On the maximum rank of a real binary form", *Annali di Matematica Pura e Applicata*, vol. 190, no. 1, pp. 55–59, 2011, arXiv:1006.5127, to appear.

[14] G. Chabriel, M. Kleinstauber, E. Moreau, H. Shen, P. Tichavsky, A. Yeredor, "Joint matrices decompositions and blind source separation", *IEEE Sig. Proc. Magazine*, vol. 31, May 2014, this issue.

[15] L. Chiantini, G. Ottaviani, "On generic identifiability of 3-tensors of small rank", *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 3, pp. 1018–1037, 2012.

[16] A. Cichocki, D. Mandic et al., "Multi-way component analysis for tensor data: a signal processing perspective", *Sig. Proc. Magazine*, 2013, to appear.

[17] P. Comon, "Tensor decompositions, state of the art and applications", in *Mathematics in Signal Processing V*, J. G. McWhirter, I. K. Proudler, Eds., pp. 1–24. Clarendon Press, Oxford, UK, 2002, arXiv:0905.0454v1.

[18] P. Comon, "Tensors, usefulness and unexpected properties", in *15th IEEE Workshop on Statistical Signal Processing (SSP'09)*, Cardiff, UK, Aug. 31 - Sep. 3 2009, pp. 781–788, keynote. hal-00417258.

[19] P. Comon, G. Golub, L-H. Lim, B. Mourrain, "Symmetric tensors and symmetric tensor rank", *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1254–1279, Sept. 2008.

[20] P. Comon, C. Jutten, Eds., *Handbook of Blind Source Separation, Independent Component Analysis and Applications*, Academic Press, Oxford UK, Burlington USA, 2010.

[21] P. Comon, X. Luciani, A. L. F. De Almeida, "Tensor decompositions, alternating least squares and other tales", *Jour. Chemometrics*, vol. 23, pp. 393–405, Aug. 2009, hal-00410057.

[22] P. Comon, G. Ottaviani, "On the typical rank of real binary forms", *Linear and Multilinear Algebra*, vol. 60, no. 6, pp. 657–667, May 2012.

[23] P. Comon, G. Ottaviani, "Tensor decompositions, a geometric viewpoint", 2014, being submitted.

[24] P. Comon, J. M. F. Ten Berge, L. De Lathauwer, J. Castaing, "Generic and typical ranks of multi-way arrays", *Lin. Alg. Appl.*, vol. 430, no. 11–12, pp. 2997–3007, June 2009, hal-00410058.

[25] L. De Lathauwer, "A link between canonical decomposition in multilinear algebra and simultaneous matrix diagonalization", *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, 2006.

[26] L. De Lathauwer, "Blind separation of exponential polynomials and the decomposition of a tensor in rank-(Lr,Lr,1) terms", *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 145–1474, 2011.

[27] L. De Lathauwer, B. De Moor, Vandewalle J., "A multilinear singular value decomposition", *SIAM Jour. Matrix Ana. Appl.*, vol. 21, no. 4, pp. 1253–1278, Apr. 2000.

[28] L. De Lathauwer, B. De Moor, Vandewalle J., "On the best rank-1 and rank-(R1,R2,. . . RN) approximation of high-order tensors", *SIAM Jour. Matrix Ana. Appl.*, vol. 21, no. 4, pp. 1324–1342, Apr. 2000.

[29] L. De Lathauwer, B. De Moor, Vandewalle J., "Independent Component Analysis and (simultaneous) third-order tensor diagonalization", *IEEE Trans. Sig. Proc.*, vol. 49, no. 10, pp. 2262–2271, Oct. 2001.

[30] L. De Lathauwer, B. De Moor, Vandewalle J., "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition", *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 2, pp. 295–327, 2004.

[31] I. Domanov, L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors - part I: Basic results and uniqueness of one factor matrix", *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 855–875, 2013.

[32] I. Domanov, L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors - part II: Overall uniqueness", *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 876–903, 2013.

[33] R. Ehrenborg, G. C. Rota, "Apolarity and canonical forms for homogeneous polynomials", *European Jour. Combinatorics*, vol. 14, no. 3, pp. 157–181, May 1993.

[34] S. Friedland, "Best rank one approximation of real symmetric tensors can be chosen symmetric", *Frontiers of Mathematics in China*, vol. 8, no. 1, pp. 19–40, 2013, special issue on Tensor Theory.

[35] S. Friedland, G. Ottaviani, *The number of singular vector tuples and uniqueness of best rank one approximation of tensors*, Oct. 2012, arXiv:1210.8316.

[36] R. A. Harshman, "Foundations of the Parafac procedure: Models and conditions for an explanatory multimodal factor analysis", *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[37] C. J. Hillar, L.-H. Lim, *Most tensor problems are NP-hard*, 2012, arXiv:0911.1393.

[38] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products", *J. Math. and Phys.*, vol. 6, no. 1, pp. 165–189, 1927.

[39] T. D. Howell, "Global properties of tensor rank", *Lin. Alg. Appl.*, vol. 22, pp. 9–23, Dec. 1978.

[40] A. Iarrobino, "Inverse system of a symbolic power. II. the Waring problem for forms", *J. Algebra*, vol. 174, no. 3, pp. 1091–1110, June 1995.

[41] T. Jiang, N. Sidiropoulos, "Kruskal's permutation lemma and the identification of Candecomp/Parafac and bilinear models", *IEEE Trans. Sig. Proc.*, vol. 52, no. 9, pp. 2625–2636, Sept. 2004.

[42] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis", *J. Chemometrics*, vol. 14, no. 3, pp. 105–122, June 2000.

[43] E. Kofidis, P. Regalia, "One the best rank-1 approximation of higher-order symmetric tensors", *SIAM Jour. Matrix Ana. Appl.*, vol. 23, no. 3, pp. 863–884, 2002.

[44] E. Kofidis, P. A. Regalia, "Tensor approximation and signal processing applications", in *Structured Matrices in Mathematics, Computer Science and Engineering*, V. Olshevsky, Ed., Providence, RI, 2001, AMS Publ.

[45] T. G. Kolda, "Orthogonal tensor decomposition", *SIAM Jour. Matrix Ana. Appl.*, vol. 23, no. 1, pp. 243–255, 2001.

[46] T. G. Kolda, B. W. Bader, "Tensor decompositions and applications", *SIAM Review*, vol. 51, no. 3, pp. 455–500, Sept. 2009.

[47] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions", *Lin. Alg. Appl.*, vol. 18, pp. 95–138, 1977.

[48] J. B. Kruskal, "Rank of N-way arrays and the geometry of 2x2x2 arrays", Tech. Rep., AT & T Bell Labs, Feb. 1985.

[49] J. M. Landsberg, *Tensors: Geometry and Applications*, vol. 128 of *Graduate Studies in Mathematics*, AMS publ., Providence, Rhode Island, 2012.

[50] S. Leurgans, R. T. Ross, R. B. Abel, "A decomposition for three-way arrays", *SIAM Jour. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 1064–1083, Oct. 1993.

[51] X. L. Li, X. D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution", *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1803–1814, 2007.

[52] T. Lickteig, "Typical tensorial rank", *Lin. Alg. Appl.*, vol. 69, pp. 95–120, 1985.

[53] L-H. Lim, "Singular values and eigenvalues of tensors: a variational approach", in *IEEE Int. Workshop on Comput. Adv. Multi-Sensor Adapt. Proc.*, Puerto Vallarta, Mexico, Dec. 13-15 2005, pp. 129–132.

[54] L-H. Lim, P. Comon, "Nonnegative approximations of nonnegative tensors", *Jour. Chemometrics*, vol. 23, pp. 432–441, Aug. 2009, hal-00410056.

[55] L.-H. Lim, P. Comon, "Blind multilinear identification", *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1260–1280, Feb. 2014, open access.

[56] X. Luciani, L. Albera, "Semi-algebraic canonical decomposition of multi-way arrays and joint eigenvalue decomposition", in *ICASSP'2011*, Prague, Mar. 12-15 2011, pp. 4104–4107, IEEE.

[57] P. Mccullagh, *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability. Chapman and Hall, 1987.

[58] R. Miranda, "Linear systems of plane curves", *Notices of the AMS*, vol. 46, no. 2, pp. 192–202, Feb. 1999.

[59] G. Ni, Y. Wang, "On the best rank-1 approximation to higher-order symmetric tensors", *Mathematical and Computer Modelling*, vol. 46, no. 9-10, pp. 1345–1352, Nov. 2007.

[60] L. Oeding, G. Ottaviani, "Eigenvectors of tensors and algorithms for Waring decomposition", *J. Symb. Comput.*, vol. 54, pp. 9–35, July 2013, arXiv:1103.0203.

[61] P. Paatero, "Construction and analysis of degenerate Parafac models", *Jour. Chemometrics*, vol. 14, no. 3, pp. 285–299, June 2000.

[62] D. T. Pham, "Joint approximate diagonalization of positive definite Hermitian matrices", *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1136–1152, 2001.

[63] A. H. Phan, A. Cichocki, P. Tichavsky, G. Luta, A. Brockmeier, "Tensor completion through multiple Kronecker product decomposition", in *ICASSP'2013*, Vancouver, May 26-31 2013, pp. 3233–3237.

[64] A. H. Phan, P. Tichavsky, A. Cichocki, "Candecomp/Parafac decomposition of high-order tensors through tensor reshaping", *IEEE Trans. Sig. Proc.*, vol. 61, pp. 4847–4860, Oct. 2013.

[65] M. Rajih, P. Comon, R. Harshman, "Enhanced line search : A novel method to accelerate Parafac", *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1148–1171, Sept. 2008.

[66] B. Reznick, "Sums of even powers of real linear forms", *Memoirs of the AMS*, vol. 96, no. 463, pp. 1–155, Mar. 1992.

[67] B. Reznick, "Laws of inertia in higher degree binary forms", *Proc. Amer. Math. Soc.*, vol. 138, no. 3, pp. 815–826, June 2010.

[68] J. A. Rhodes, "A concise proof of kruskal's theorem on tensor decomposition", *Lin. Alg. Appl.*, vol. 432, no. 7, pp. 1818–1824, Mar. 2010.

[69] F. Roemer, M. Haardt, "A semi-algebraic framework for approximate CP decompositions via simultaneous matrix diagonalizations (secsi)", *Signal Processing*, vol. 93, no. 9, pp. 2722–2738, Sept. 2013.

[70] J-P. Royer, N. Thirion-Moreau, P. Comon, "Computing the polyadic decomposition of nonnegative third order tensors", *Signal Processing*, vol. 91, no. 9, pp. 2159–2171, Sept. 2011, hal-00618729.

[71] J. R. Ruiz-Tolosa, E. Castillo, *From Vectors to Tensors*, Universitext. Springer, Berlin, Heidelberg, 2005.

[72] A. Schönhage, "Partial and total matrix multiplication", *SIAM J. Computation*, vol. 10, no. 3, pp. 434–455, 1981.

[73] N. D. Sidiropoulos, R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays", *Jour. Chemo.*, vol. 14, no. 3, pp. 229–239, June 2000.

[74] N. D. Sidiropoulos, R. Bro, G. B. Giannakis, "Parallel factor analysis in sensor array processing", *IEEE Trans. Sig. Proc.*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.

[75] V. De Silva, L-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem", *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1084–1127, 2008.

[76] A. Smilde, R. Bro, P. Geladi, *Multi-Way Analysis*, Wiley, Chichester UK, 2004.

[77] L. Sorber, M. Van Barel, L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: canonical polyadic decomposition, decomposition in rank-s terms and a new generalization", *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 695–720, Apr. 2013.

[78] M. Sorensen, P. Comon, "Tensor decompositions with banded matrix factors", *Lin. Alg. Appl.*, vol. 438, no. 1, pp. 919–941, Jan. 2013.

[79] A. Stegeman, , L. De Lathauwer, "A method to avoid diverging components in the Candecomp/Parafac model for generic IxJx2 arrays", *SIAM J. Matrix Ana. Appl.*, vol. 30, no. 4, pp. 1614–1638, 2009.

[80] A. Stegeman, P. Comon, "Subtracting a best rank-1 approximation does not necessarily decrease tensor rank", *Lin. Alg. Appl.*, vol. 433, no. 7, pp. 1276–1300, Dec. 2010, hal-00512275.

[81] A. Stegeman, N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition", *Lin. Alg. Appl.*, vol. 420, no. 2-3, pp. 540–552, Jan. 2007.

[82] V. Strassen, "Rank and optimal computation of generic tensors", *Lin. Alg. Appl.*, vol. 52, pp. 645–685, July 1983.

[83] P. Tichavsky, Z. Koldovsky, "Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources", *IEEE Trans. Sig. Proc.*, vol. 59, no. 3, pp. 1037–1047, Mar. 2011.

[84] G. Tomasi, R. Bro, "A comparison of algorithms for fitting the Parafac model", *Comp. Stat. Data Anal.*, vol. 50, pp. 1700–1734, 2006.

[85] L. R. Tucker, "Some mathematical notes for three-mode factor analysis", *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sept. 1966.

[86] R. Vollgraf, K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization", *IEEE Trans. Sig,. Proc*, vol. 54, pp. 3270–3278, Sept. 2006.

[87] A. Yeredor, "Non-orthogonal joint diagoinalization in the LS sense with application in blind source separation", *IEEE Trans. Sig. Proc.*, vol. 50, no. 7, pp. 1545–1553, 2002.

[88] X. Zhang, C. Ling, L. Qi, "The best rank-1 approximation of a symmetric tensor and related spherical optimization problems", *SIAM J. Matrix Analysis Appl.*, vol. 33, no. 3, pp. 806–821, 2012.

[89] A. Ziehe, G. Nolte, K. R. Müller, "A fast algorithm for joint diagonalization with non orthogonal transformations and its application to blind source separation", *J. Machine Learning Research*, vol. 5, pp. 777–800, Dec. 2004.

[90] R. Bro. http://www.models.kvl.dk/users/rasmus

[91] L. De Lathauwer. http://homes.esat.kuleuven.be/ delathau

[92] T. Kolda. http://www.sandia.gov/ tgkolda

[93] A. H. Phan. http://www.bsp.brain.riken.jp/ phan

[94] P. Comon. http://www.gipsa-lab.grenoble-inp.fr/ pierre.comon

[95] P. Kroonenberg. Three-Mode Company. http://three-mode.leidenuniv.nl